Distributional Effects of a Continuous Treatment with an Application on Intergenerational Mobility

Brantly Callaway* Weige Huang†

Supplementary Appendix

July 27, 2018

This supplementary appendix contains additional results for the paper "Distributional Effects of a Continuous Treatment with an Application on Intergenerational Mobility." The first set of results show that all the parameters of interest are indeed Hadamard differentiable maps of $F_{Y|T}$ and $F^C_{Y|T}$ and therefore the results from Corollary 1 hold. This part also provides explicit expressions for each of the terms in Corollary 1 for each parameter of interest. The second set of results includes additional details and results for estimating the conditional distribution $F_{Y|T,X}(y|t,x)$ using distribution regression and compares the results in the main part of the paper, which use quantile regression, to results using distribution regression in the first step.

1 Additional Details on Asymptotic Results

This section shows that all the parameters that we consider are Hadamard differentiable maps of $F_{Y|T}$ and $F^C_{Y|T}$, the validity of using the empirical bootstrap for inference, and details on how to test if a particular parameter changes with the treatment level.

1.1 Hadamard Differentiability of Parameters of Interest

The first result establishes the limiting process for the fraction of individuals who have "low" outcomes (e.g. child's income being below the poverty line) as a function of the treatment.

SA Theorem 1. *Let $\hat{G}^{POV}_T(y_p|t) = \sqrt{n}(\hat{F}_{Y|T}(y_p|t) - F_{Y|T}(y_p|t))$ where $y_p \in \mathcal{Y}$ denotes a particular value for a "low" outcome (e.g. the poverty line) and is fixed; let $\hat{G}^{C,POV}_T(y_p|t) = \sqrt{n}(\hat{F}^C_{Y|T}(y_p|t) - F^C_{Y|T}(y_p|t))$. Under Assumptions 2 to 5 and Assumption A.1,*

$$(\hat{G}^{POV}_T(y_p|t), \hat{G}^{C,POV}_T(y_p|t)) \rightsquigarrow (\mathbb{V}^{POV}_{y_p}(t), \mathbb{V}^{C,POV}_{y_p}(t))$$

where $\mathbb{V}^{POV}(t)$ is a stochastic process in the metric space $l^\infty(\mathcal{T})$ given by $\mathbb{V}_{Y|T}(y_p|t)$ and where $\mathbb{V}^{C,POV}(t)$ is a stochastic process in the metric space $l^\infty(\mathcal{T})$ given by $\mathbb{V}^C_{Y|T}(y_p|t)$. In addition,

$$\sqrt{n}(\hat{\Delta}^{POV}(y_p, t) - \Delta^{POV}(y_p, t)) \rightsquigarrow \mathbb{V}^{POV}_{y_p}(t) - \mathbb{V}^{C,POV}_{y_p}(t)$$

*Assistant Professor, Department of Economics, Temple University. Email: brantly.callaway@temple.edu
†Corresponding Author, PhD Student, Department of Economics, Temple University. Email: weige.huang@temple.edu

in the space $l^\infty(\mathcal{T})$.

Proof. The result follows immediately from Theorem 1. □

The results of SA Theorem 1 apply to the case where the fraction of individuals with a "low" outcome is computed using the observed conditional distribution, the counterfactual distribution, and also provides the limiting process for their difference. This last result allows one to formally test whether accounting for differences in covariates across different levels of the treatment accounts for differences in the fraction of individuals with "low" outcomes.

The next result shows that the observed quantiles of the outcome conditional on the treatment and the counterfactual quantiles of the outcome conditional on the treatment converge uniformly both in the quantile τ and the treatment to a Gaussian process.

SA Theorem 2. *Let* $\hat{Z}_{Y|T}(\tau|t) = \sqrt{n}(\hat{Q}_{Y|T}(\tau|t) - Q_{Y|T}(\tau|t))$ *and let* $\hat{Z}^C_{Y|T}(\tau|t) = \sqrt{n}(\hat{Q}^C_{Y|T}(\tau|t) - Q^C_{Y|T}(\tau|t))$. *Let* $\mathbb{S} = l^\infty(\mathcal{UT})^2$. *Under Assumptions 2 to 5 and Assumption A.1,*

$$(\hat{Z}_{Y|T}(\tau|t), \hat{Z}^C_{Y|T}(\tau|t)) \rightsquigarrow (\mathbb{Z}(\tau|t), \mathbb{Z}^C(\tau|t))$$

in the space $\mathbb{S}$ *with*

$$\mathbb{Z}(\tau|t) = \frac{\mathbb{V}_{Y|T}(Q_{Y|T}(\tau|t))}{f_{Y|T}(Q_{Y|T}(\tau|t)|t)} \qquad \text{and} \qquad \mathbb{Z}^C(\tau|t) = \frac{\mathbb{V}^C_{Y|T}(Q^C_{Y|T}(\tau|t))}{f^C_{Y|T}(Q^C_{Y|T}(\tau|t)|t)}$$

Moreover,

$$\sqrt{n}(\hat{\Delta}^Q(\tau|t) - \Delta^Q(\tau|t)) \rightsquigarrow \mathbb{Z}(\tau|t) - \mathbb{Z}^C(\tau|t)$$

in the space $l^\infty(\mathcal{T})$.

Proof. The result follows from SA Theorem 1 and by Lemma 3.9.23(ii) in Van Der Vaart and Wellner (1996). □

Quantiles of counterfactual distributions may be of interest in themselves. We also use SA Theorem 2 in the next set of theorems. SA Theorem 2 is also useful as a building block for other parameters of interest, in particular for $E[Y|T = t]$, $Var(Y|T = t)$, and $IQR(\tau_1, \tau_2; t)$ as well as their counterfactual counterparts.

The next theorem establishes the limiting process for the average outcome as a function of the treatment which holds uniformly in the treatment. Here we define the average as the trimmed version of the average, i.e. $E[Y|T = t] = \int_\epsilon^{1-\epsilon} Q_{Y|T}(\tau|t)\, d\tau$ and $E^C[Y|T = t] = \int_\epsilon^{1-\epsilon} Q^C_{Y|T}(\tau|t)\, d\tau$, but noting that under some additional conditions one may be able to integrate all the way to 0 and 1 (see Footnote 6 in the main text).

SA Theorem 3. *Let* $\hat{G}^E_T(t) = \sqrt{n}(\hat{E}[Y|T = t] - E[Y|T = t])$ *and let* $\hat{G}^{C,E}_T(t) = \sqrt{n}(\hat{E}^C[Y|T = t] - E^C[Y|T = t])$. *Under Assumptions 2 to 5 and Assumption A.1,*

$$(\hat{G}^E_T(t), \hat{G}^{C,E}_T(t)) \rightsquigarrow (\mathbb{V}^E_T(t), \mathbb{V}^{C,E}_T(t))$$

in the space $l^\infty(\mathcal{T})^2$ *where* $\mathbb{V}^E_T$ *is a tight Gaussian process with mean 0 given by* $\mathbb{V}^E_T = \int_\epsilon^{1-\epsilon} \mathbb{Z}(\tau|\cdot)\, \mathrm{d}\tau$ *and* $\mathbb{V}^E_T$ *is a tight Gaussian process with mean 0 given by* $\mathbb{V}^{C,E}_T = \int_\epsilon^{1-\epsilon} \mathbb{Z}^C(\tau|\cdot)\, \mathrm{d}\tau$. *In addition,*

$$\sqrt{n}(\hat{\Delta}^E(t) - \Delta^E(t)) \rightsquigarrow \mathbb{V}^E_T(t) - \mathbb{V}^{C,E}_T(t)$$

in the space $l^\infty(\mathcal{T})$.

Before proving SA Theorem 3, we first prove the following lemma.

SA Lemma 1. *Let* $\mathbb{D} = l^\infty(\mathcal{UT})$ *and define the map* $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto l^\infty(\mathcal{T})$ *given by*

$$\phi(\Phi) = \int_\epsilon^{1-\epsilon} \Phi(\tau|\cdot)\mathrm{d}\tau$$

for $\Phi \in \mathbb{D}$*. Then, the map* ϕ *is Hadamard differentiable at* Φ_0 *tangentially to* $\mathbb{D}$ *with derivative at* Φ_0 *in* $\xi \in \mathbb{D}$ *given by*

$$\phi'_{\Gamma_0}(\xi) = \int_\epsilon^{1-\epsilon} \xi(\tau|\cdot)\mathrm{d}\tau$$

Proof. Consider any sequence $t_k > 0$ and $\Phi_k \in \mathbb{D}$ for $k = 1, 2, 3, \ldots$ with $t_k \downarrow 0$ and

$$\xi_k = \frac{\Phi_k - \Phi}{t_k} \to \xi \in \mathbb{D} \text{ as } k \to \infty$$

Then,

$$\begin{aligned}
\frac{\phi(\Phi_k) - \phi(\Phi)}{t_k} - \phi'_\Phi(\xi) &= \int_\epsilon^{1-\epsilon} \frac{\Phi_k(\tau|\cdot) - \Phi(\tau|\cdot)}{t_k} \,\mathrm{d}\tau - \int_\epsilon^{1-\epsilon} \xi(\tau|\cdot)\,\mathrm{d}\tau \\
&= \int_\epsilon^{1-\epsilon} \xi_k(\tau|\cdot) - \xi(\tau|\cdot)\,\mathrm{d}\tau \\
&\leq \|\xi_k - \xi\|_\infty \int_\epsilon^{1-\epsilon} \mathrm{d}\tau \to 0 \text{ as } k \to \infty
\end{aligned}$$

□

Proof of SA Theorem 3. For the first part,

$$\begin{aligned}
\sqrt{n}(\hat{E}[Y|T=t] - E[Y|T=t]) &= \sqrt{n}\left(\frac{1}{S}\sum_{s=1}^{S} \hat{Q}_{Y|T}(\tau_s|t) - \phi(\hat{Q}_{Y|T})\right) + \sqrt{n}(\phi(\hat{Q}_{Y|T}) - \phi(Q_{Y|T})) \\
&= \phi'_{Q_{Y|T}}\sqrt{n}(\hat{Q}_{Y|T} + Q_{Y|T}) + o_p(1)
\end{aligned}$$

which holds uniformly in t under the condition that S is large enough, e.g. $S = Cn^{1/2+\delta}$ for some $C > 0$ and $\delta > 0$, and by SA Lemma 1. This implies the result.

The second and third parts follows exactly the same argument. □

SA Theorem 3 provides a way to construct uniform confidence bands for $E[Y|T=t]$, $E^C[Y|T=t]$, and their difference. In the application on intergenerational mobilty, these results allow us to compare average child's income across parents' income and learn about the role that covariates play in the intergenerational transmission of income.

The next result establishes the limiting process for the variance of the outcome conditional on the treatment; here, like for the mean, we consider a trimmed version of the variance.

SA Theorem 4. *Under Assumptions 2 to 5 and Assumption A.1,*

$$\sqrt{n}(\hat{Var}(Y|T=t) - Var(Y|T=t)) \rightsquigarrow \mathbb{V}_T^V(t)$$

in the space $l^\infty(\mathcal{T})$ *where* $\mathbb{V}_T^V$ *is tight Gaussian process with mean 0 that is given by*

$$\mathbb{V}_T^V(t) = 2\int_\epsilon^{1-\epsilon} (Q_{Y|T}(\tau|t) - E[Y|T=t]) \left(\mathbb{Z}(\tau,t) - \int_\epsilon^{1-\epsilon} \mathbb{Z}(u,t)\ \mathrm{d}u\right)\ \mathrm{d}\tau$$

In addition,

$$\sqrt{n}(\hat{Var}^C(Y|T=t) - Var^C(Y|T=t)) \rightsquigarrow \mathbb{V}_T^{C,V}(t)$$

in the space $l^\infty(\mathcal{T})$ *where* $\mathbb{V}_T^{C,V}$ *is a tight Gaussian process with mean 0 that is given by*

$$\mathbb{V}_T^{C,V}(t) = 2\int_\epsilon^{1-\epsilon} (Q_{Y|T}^C(\tau|t) - E^C[Y|T=t]) \left(\mathbb{Z}^C(\tau,t) - \int_\epsilon^{1-\epsilon} \mathbb{Z}^C(u,t)\ \mathrm{d}u\right)\ \mathrm{d}\tau$$

Finally,

$$\sqrt{n}(\hat{\Delta}^{Var}(t) - \Delta^{Var}(t)) \rightsquigarrow \mathbb{V}_T^{\Delta,Var}(t)$$

where $\mathbb{V}_T^{\Delta,Var}$ *is a tight Gaussian process with mean 0 that is given by* $\mathbb{V}_T^{\Delta,Var} = \mathbb{V}_T^V - \mathbb{V}_T^{C,V}$.

Before proving the theorem, we prove the following lemma.

SA Lemma 2. *Let* $\mathbb{D} = l^\infty(\mathcal{UT})$ *and define the map* $\pi : \mathbb{D}_\pi \subset \mathbb{D} \mapsto l^\infty(\mathcal{T})$ *by*

$$\pi(\Gamma) = \int_\epsilon^{1-\epsilon} \left(\Gamma(\tau|\cdot) - \int_\epsilon^{1-\epsilon} \Gamma(u|\cdot)\ \mathrm{d}u\right)^2\ \mathrm{d}\tau$$

for $\Gamma \in \mathbb{D}$. *Then, the map* π *is Hadamard differentiable at* Γ_0 *tangentially to* $\mathbb{D}$ *with derivative at* Γ_0 *in* $\gamma \in \mathbb{D}$ *given by*

$$\pi'_{\Gamma_0}(\gamma) = 2\int_\epsilon^{1-\epsilon} \left\{\left(\Gamma_0(\tau|\cdot) - \int_\epsilon^{1-\epsilon} \Gamma_0(u|\cdot)\ \mathrm{d}u\right)\left(\gamma(\tau|\cdot) - \int_\epsilon^{1-\epsilon} \gamma(u|\cdot)\ \mathrm{d}u\right)\right\}\ \mathrm{d}\tau$$

Proof. Consider the maps $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto l^\infty(\mathcal{T})$ given in SA Lemma 1, $\pi_1 : \mathbb{D}_{\pi_1} \subset l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T}) \mapsto \mathbb{D}_{\pi_2}$ given by

$$\pi_1(\Lambda) = \Lambda_1 - \Lambda_2$$

for $\Lambda = (\Lambda_1, \Lambda_2) \in l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T})$ and the map $\pi_2 : \mathbb{D}_{\pi_2} \subset l^\infty(\mathcal{UT}) \mapsto \mathbb{D}_\phi$ given by

$$\pi_2(\Theta) = \Theta^2$$

for $\Theta \in l^\infty(\mathcal{UT})$

First, notice that the map π is given by the composition map $\pi(\Gamma) = \phi \circ \pi_2 \circ \pi_1(\Gamma, \phi(\Gamma))$. Lemma 3.9.3 of Van Der Vaart and Wellner (1996) implies

$$\pi'_{\Gamma_0}(\gamma) = \phi'_{\pi_2\circ\pi_1(\Gamma_0,\phi(\Gamma_0))} \circ \pi'_{2,\pi_1(\Gamma_0,\phi(\Gamma_0))} \circ \pi'_{1,(\Gamma_0,\phi(\Gamma_0))}(\gamma, \phi'_{\Gamma_0}(\gamma)) \tag{1.1}$$

Next, the map π_1 is Hadamard differentiable at $\Lambda_0 = (\Lambda_{10}, \Lambda_{20}) \in l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T})$ tangentially to $l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T})$ with derivative at Λ_0 in $\lambda = (\lambda_1, \lambda_2)$ in $l^\infty(\mathcal{UT}) \times l^\infty(\mathcal{T})$ given by

$$\pi'_{1,\Lambda_0}(\lambda) = \lambda_1 - \lambda_2 \tag{1.2}$$

Next, the map π_2 is Hadamard differentiable at Θ_0 tangentially to $l^\infty(\mathcal{UT})$ with derivative at Θ_0 in $\theta \in l^\infty(\mathcal{UT})$ given by

$$\pi'_{2,\Theta_0}(\theta) = 2\Theta_0\theta \tag{1.3}$$

Consider any sequence $t_k > 0$ and $\Theta_k \in l^\infty(\mathcal{UT})$ and for $k = 1, 2, 3, \ldots$ $t_k \downarrow 0$ and

$$\theta_k = \frac{\Theta_k - \Theta_0}{t_k} \to \theta \in l^\infty(\mathcal{UT}) \text{ as } k \to \infty$$

Then,

$$\begin{aligned}\frac{\pi_2(\Theta_k) - \pi_2(\Theta_0)}{t_k} - \pi'_{2,\Theta_0}(\theta) &= \frac{(\Theta_0 + t_k\theta_k)^2 - \Theta_0^2}{t_k} - 2\Theta_0\theta \\ &= 2\Theta_0(\theta_k - \theta) + \theta_k^2 t_k \to 0 \text{ as } k \to \infty\end{aligned}$$

which shows the result. And the main result follows from SA Lemma 2 and Equations 1.1, 1.2, and 1.3.

□

Proof of SA Theorem 4. For the first part of the result,

$$\begin{aligned}\sqrt{n}\left(\hat{Var}(Y|T=t) - Var(Y|T=t)\right) &= \sqrt{n}\left(\frac{1}{S}\sum_{s=1}^{S}\left(\hat{Q}_{Y|T}(\tau_s|t) - \hat{E}[Y|T=t]\right)^2 - \pi(\hat{Q}_{Y|T})\right) \\ &\quad + \sqrt{n}(\pi(\hat{Q}_{Y|T}) - \pi(Q_{Y|T})) \\ &= \pi'_{Q_{Y|T}}\sqrt{n}(\hat{Q}_{Y|T} - Q_{Y|T}) + o_p(1)\end{aligned}$$

which holds uniformly in t as long as S is large enough, e.g. $S = Cn^{1/2+\delta}$ for $C > 0$ and $\delta > 0$, and by SA Lemma 3. This implies the result. The second and third parts of the result hold using the same arguments. □

The final theorem in this section provides the limiting process of the inter-quantile range of the outcome conditional on the treatment.

SA Theorem 5. *Under Assumptions 2 to 5 and Assumption A.1,*

$$\sqrt{n}(I\hat{Q}R(\tau_1, \tau_2, t) - IQR(\tau_1, \tau_2, t)) \rightsquigarrow \mathbb{G}_T^{IQR}(\tau_1, \tau_2, t)$$

in the space $l^\infty(\mathcal{T})$ where $\mathbb{G}_T^{IQR}$ is a tight mean 0 Gaussian process given by $\mathbb{G}_T^{IQR}(\tau_1, \tau_2, t) = \mathbb{Z}(\tau_1|t) - \mathbb{Z}(\tau_2|t)$ where $\mathbb{Z}$ is given in SA Theorem 2. Also,

$$\sqrt{n}(I\hat{Q}R^C(\tau_1, \tau_2, t) - IQR^C(\tau_1, \tau_2, t)) \rightsquigarrow \mathbb{G}_T^{C,IQR}(\tau_1, \tau_2, t)$$

in the space $l^\infty(\mathcal{T})$ where $\mathbb{G}_T^{C,IQR}$ is a tight Gaussian process with mean 0 given by $\mathbb{G}_T^{C,IQR}(\tau_1, \tau_2, t) = \mathbb{Z}^C(\tau_1|t) - \mathbb{Z}^C(\tau_2|t)$ where $\mathbb{Z}^C$ is given in SA Theorem 2. Finally,

$$\sqrt{n}(\hat{\Delta}^{IQR}(\tau_1, \tau_2, t) - \Delta^{IQR}(\tau_1, \tau_2, t)) \rightsquigarrow \mathbb{G}_T^{IQR}(\tau_1, \tau_2, t) - \mathbb{G}_T^{C,IQR}(\tau_1, \tau_2, t)$$

Proof. The result follows immediately from SA Theorem 2 □

1.2 Inference using the Bootstrap

The limiting processes above depend on unknown nuisance parameters which complicate inference. Thus, to conduct inference, we use the empirical bootstrap. This section shows that the empirical bootstrap procedure can be used to construct asymptotically valid uniform bands for each of the parameters considered above.

Let $\hat{\theta}^*(t)$ denote a bootstrapped version of the estimator; in other words, computed using draws from the empirical distribution $\hat{F}_{Y,T,X}$ in the same manner as $\hat{\theta}(t)$.

SA Theorem 6. *Under Assumptions 2 to 5 and Assumption A.1,*

$$\sqrt{n}(\hat{\theta}^*(t) - \hat{\theta}(t)) \rightsquigarrow_* \mathbb{V}_{\theta T}(t)$$

where $\rightsquigarrow_*$ indicates weak convergence under the bootstrap law and $\mathbb{V}_{\theta T}$ is the tight mean 0 Gaussian process for each parameter $\theta(t)$ given above.

As a first step, we prove the following lemma.

SA Lemma 3. *Under Assumptions 2 to 5 and Assumption A.1,*

$$(\hat{G}^*_{Y|T}(y|t), \hat{G}^{C*}_{Y|T}(y|t)) \rightsquigarrow_* (\mathbb{V}_{Y|T}, \mathbb{V}^C_{Y|T})$$

where $\rightsquigarrow_*$ indicates weak convergence under the bootstrap law and $(\mathbb{V}_{Y|T}, \mathbb{V}^C_{Y|T})$ is the Gaussian process from SA Theorem 1.

Proof. The result follows from SA Theorem 1 and by Theorem 3.6.1 of Van Der Vaart and Wellner (1996). □

Proof of SA Theorem 6. The result follows from SA Lemma 3, that $\theta(t)$ is a Hadamard differentiable function of the $F_{Y|T}$ and $F^C_{Y|T}$, and by the functional delta method applied to the bootstrap. □

1.3 Testing if parameters depend on the value of the treatment

We are also interested in testing whether each of the parameters of interest depends on t as discussed in Section 2.3. As in the previous section, let $\theta(t)$ generically denote one of the parameters of interest. The results in SA Theorems 1 to 5 imply that $\sqrt{n}(\hat{\theta}(t) - \theta(t)) \rightsquigarrow \mathbb{V}_{\theta T}$ in the space $l^\infty(\mathcal{T})$ where $\hat{\theta}(t)$ is the estimator of $\theta(t)$ and $\mathbb{V}_{\theta T}$ is some tight mean 0 Gaussian process that depends on which parameter is being estimated. Each of the parameters considered in the paper satisfies the following condition.

Condition 1. *Denote any of the parameters considered above by $\theta(t)$ and its estimator given in the parameter by $\hat{\theta}(t)$. Also, for $\mu_\theta = E[\theta(T)]$, let*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_{\theta i} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta(T_i) - \mu_\theta)$$

denote the influence function for estimating μ_θ when θ is known. Then,

$$\left(\sqrt{n}(\hat{\theta}(t) - \theta(t)), \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_{\theta i} \right) \rightsquigarrow (\mathbb{V}_{\theta T}(t), \mathbb{W}_\theta)$$

in the space $l^\infty(\mathcal{T})$.

Consider $R_\theta(t) = \theta(t) - E[\theta(T)]$. In this section, we are interested in forming uniform confidence bands for $R_\theta(t)$ as well as testing the null hypothesis

$$H_0 : R_\theta(t) = 0 \qquad \text{for all } t \in \mathcal{T} \tag{1.4}$$

A natural estimator of $R_\theta(t)$ is given by $\hat{R}_\theta(t) = \hat{\theta}(t) - \frac{1}{n}\sum_{i=1}^n \hat{\theta}(T_i)$. The next result establishes the limiting process for $\hat{R}_\theta(t)$.

Proposition 1. *Under Assumptions 2 to 5 and Assumption A.1 and Condition 1*

$$\sqrt{n}\left(\hat{R}_\theta(t) - R_\theta(t)\right) \rightsquigarrow V^R_{\theta T}(t)$$

in the space $l^\infty(\mathcal{T})$ where $\mathbb{V}^R_{\theta T}(t)$ is a tight mean 0 Gaussian process given by $\mathbb{V}^R_{\theta T}(t) = \mathbb{V}_{\theta T}(t) + \int_{\mathcal{T}} \mathbb{V}_{\theta T}(t)\, \mathrm{d}F_T(t) + \mathbb{W}_\theta$.

Before proving Proposition 1, we prove the following lemma.

SA Lemma 4. *Consider the map $\psi : \mathbb{D}_\psi \subset l^\infty(\mathcal{T}) \mapsto \mathbb{R}$ given by*

$$\psi(\Lambda) = \int_{\mathcal{T}} \Lambda(t)\ dF_T(t)$$

in $\Lambda \in l^\infty(\mathcal{T})$. Then, the map ψ is Hadamard differentiable at Λ_0 tangentially to $l^\infty(\mathcal{T})$ with derivative at Λ_0 in $\lambda \in l^\infty(\mathcal{T})$ given by

$$\psi'_{\Lambda_0}(\lambda) = \int_{\mathcal{T}} \lambda(t)\ dF_T(t)$$

Proof. The proof of this result follows using essentially the same arguments as in SA Lemma 1, though this case is somewhat easier. □

Proof of Proposition 1.

$$\begin{aligned}
\sqrt{n}\left(\hat{R}_\theta(t) - R_\theta(t)\right) &= \sqrt{n}\left(\hat{\theta}(t) - \int_{\mathcal{T}} \hat{\theta}(t)\ d\hat{F}_T(t)\right) - \sqrt{n}\left(\theta(t) - \int_{\mathcal{T}} \theta(t)\ dF_T(t)\right) \\
&= \sqrt{n}(\hat{\theta}(t) - \theta(t)) && (1.5) \\
&\quad + \sqrt{n}\int_{\mathcal{T}} \hat{\theta}(t) - \theta(t)\ d\hat{F}_T(t) && (1.6) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \theta(T_i) - E[\theta(T)] && (1.7)
\end{aligned}$$

The term in Equation 1.5 weakly converges to $\mathbb{V}_{\theta T}$ and the term in Equation 1.7 weakly converges to $\mathbb{W}$ – both of these hold by Condition 1.

For Equation 1.6, note that

$$\begin{aligned}
\sqrt{n}\int_{\mathcal{T}} \hat{\theta}(t) - \theta(t)\ d\hat{F}_T(t) &= \sqrt{n}\int_{\mathcal{T}} \hat{\theta}(t) - \theta(t)\ d(\hat{F}_T - F_T)(t) + \sqrt{n}\int_{\mathcal{T}} \hat{\theta}(t) - \theta(t)\ dF_T(t) \\
&= \sqrt{n}\int_{\mathcal{T}} \hat{\theta}(t) - \theta(t)\ dF_T(t) + o_p(1)
\end{aligned}$$

which weakly converges to $\int_{\mathcal{T}} V_{\theta T}(t)\ dF_T(t)$ by SA Lemma 4. This implies the result. □

Proposition 1 can be used as the basis for constructing uniform confidence bands that asymptotically cover the entire curve with $(1-\alpha)$ probability. To do this, we use the empirical bootstrap. Let $\hat{R}^*_\theta(t)$ denote the bootstrap version of $R_\theta(t)$. Given the result in Proposition 1, the following result follows

Corollary 1. *Under Assumptions 2 to 5 and Assumption A.1 and under Condition 1*

$$\sqrt{n}\left(\hat{R}^*_\theta(t) - \hat{R}_\theta(t)\right) \rightsquigarrow_* V^R_{\theta T}(t)$$

where $V^R_{\theta T}(t)$ is the Gaussian process given in Proposition 1.

The next corollary shows how to test H_0 given in Equation 1.4

Corollary 2. *Let $KS_\theta = \sup_{t\in\mathcal{T}} \Sigma_R(t)^{-1/2}|R_\theta(t)|$ and $\hat{KS}_\theta = \sup_{t\in\mathcal{T}} \hat{\Sigma}_R(t)^{-1/2}|\hat{R}_\theta(t)|$. Here, $\Sigma_R(t)$ is the asymptotic variance function of $\sqrt{n}(\hat{R}_\theta(t)-R_\theta(t))$ and $\hat{\Sigma}_R(t)$ is a uniformly consistent estimate of $\Sigma_R(t)$. Then, under H_0 (and under Assumptions 2 to 5 and Assumption A.1 and Condition 1),*

$$\sqrt{n}\left(\hat{KS}_\theta - KS_\theta\right) \rightsquigarrow \sup_{t\in\mathcal{T}} \Sigma_R(t)^{-1/2}|\mathbb{V}^R_{\theta T}|$$

Moreover, let $\hat{KS}^*_\theta$ denote the bootstrapped version of KS_θ. Then,

$$\sqrt{n}\left(\hat{KS}^*_\theta - \hat{KS}_\theta\right) \rightsquigarrow_* \sup_{t\in\mathcal{T}} \Sigma_R(t)^{-1/2}|\mathbb{V}^R_{\theta T}|$$

Corollary 2 follows immediately from the continuous mapping theorem. It shows that one can test H_0 by comparing $\hat{KS}_\theta$ to a critical value given by the $(1-\alpha)$ quantile of the bootstrapped $\sqrt{n}\left(\hat{KS}^*_\theta - \hat{KS}_\theta\right)$ which can be simulated a large number of times.

The last corollary of this section shows how to construct uniformly valid confidence bands for $R_\theta(t)$.

Corollary 3. *Under Assumptions 2 to 5 and Assumption A.1 and Condition 1 and consider the confidence region given by*

$$\hat{C}^R_\theta(t) = \hat{R}_\theta(t) \pm \hat{c}^R_{1-\alpha}\hat{\Sigma}_R(t)^{1/2}/\sqrt{n}$$

where $\hat{c}^R_{1-\alpha}$ is the $(1-\alpha)$ quantile of $\sqrt{n}\left(\hat{R}^*_\theta(t) - \hat{R}_\theta(t)\right)$ which can be simulated a large number of times and where $\hat{\Sigma}_R(t)$ is the same as in Corollary 2. Then,

$$\lim_{n\to\infty} P(R_\theta(t) \in \hat{C}^\theta_R(t) \textit{ for all } t \in \mathcal{T}) = 1-\alpha$$

2 Alternative Specifications for $F_{Y|T,X}$

In this section, we consider estimating the conditional distribution $F_{Y|T,X}$ using distribution regression rather than quantile regression. The main idea here is to estimate a series of binary response models using $\mathbb{1}\{Y \leq y\}$ as the dependent variable while varying y; that is

$$\begin{aligned} F_{Y|T,X}(y|t,x) =& E[\mathbf{1}\{Y \leq y\}|T=t, X=x] \\ =& \Lambda(\alpha_1(y)t + x'\alpha_2(y)) \end{aligned}$$

where Λ is a known link function – we use the logistic link function though one could make some other choice here. $\alpha_1(y)$ and $\alpha_2(y)$ are unknown parameters corresponding to each y, i.e., the parameters α_1 and α_2 change as y changes. $\mathbb{1}\{Y \leq y\}$ is an indicator function that equals one if $Y \leq y$ is true and zero otherwise.

To implement the distribution regression estimator, we estimate a series of logit models over a fine grid of possible values for y. The estimated conditional distribution is[1]

$$\hat{F}_{Y|T,X}(y|t,x) = \Lambda(\hat{\alpha}_1(y)t + x'\hat{\alpha}_2(y)) \tag{2.1}$$

We plug these estimators in to the counterfactual operations discussed next.

Similarly, we also estimate $F_{Y|T}(y|t)$ (the observed distribution of the outcome conditional on the treatment) using distribution regression. Here, we suppose that

$$F_{Y|T}(y|t) = \Lambda(\beta_0(y) + \beta_1(y)t)$$

and estimate the parameters $\beta_0(y)$ and $\beta_1(y)$ using logit over a fine grid of values for y. Then, the estimated value of $F_{Y|T}(y|t)$ is given by

$$\hat{F}_{Y|T}(y|t) = \Lambda(\hat{\beta}_0(y) + \hat{\beta}_1(y)t)$$

The other remaining steps of our procedure are exactly the same as in the case where a researcher uses quantile regression as in the main text of the paper.

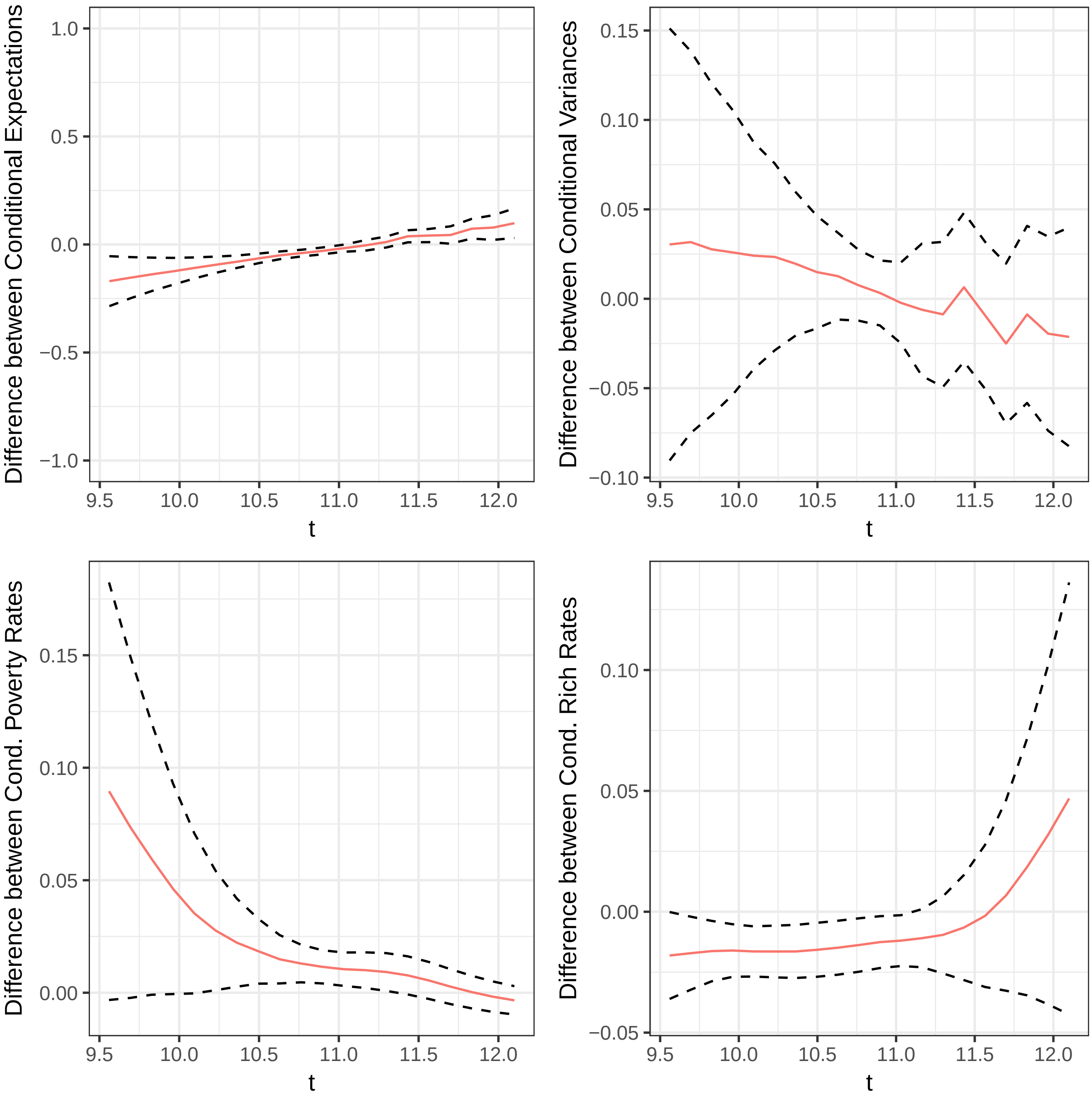
2.1 Intergenerational Mobility Results using Distribution Regression

In the final part of this Supplementary Appendix, we show SA Figure 1 to compare our main results using QR and DR for the counterfactual distribution. For the mean of child's income as a function of parents' income, the DR estimates tend to be somewhat steeper, indicating a higher intergenerational elasticity and somewhat mitigating the role of covariates (though not entirely). The DR and QR estimates of average child's income as a function of parents' income are statistically different for relatively low values of parents' income and for some high values of parents' income. The results coming from DR tend to have lower estimates of average child's income for children whose parents had relatively low income, and the DR estimates tend to have higher estimates of average child's income for children whose parents had relatively high income. Differences between estimates of the variance are not statistically different at any value of parents' income and the differences are quantitatively small. Differences in the fraction below the poverty line are only statistically different from 0 for middle values of parents' income (though the differences are quantitatively small) although using DR tends to increase the fraction of individuals estimated to be below the poverty line for low values of parents' income. For the fraction of "rich" children, the estimates are similar for DR and QR though they are statistically different for some middle values of parents' income.

Using DR instead of QR does change the magnitude of some of our estimates, but it does not change any of the qualitative results in the paper.

[1] Since the estimated conditional distribution obtained above may be non-monotonic in y, we apply the monotonization method of Chernozhukov, Fernández-Val, and Galichon 2010 based on rearrangement.

Figure 1: Differences between QR and DR estimates



Notes: This figure contains differences in estimated counterfactual parameters of interest using distribution regression and quantile regression. The plots contain the estimated parameter of interest using DR minus the estimated parameter of interest using QR along with a uniform confidence band. The top left panel plots the difference in average child's income as a function of parents' income. The top right panel plots the difference in the variance of child's income as a function of parents' income. The bottom left panel plots the difference in the fraction of children with income below the poverty line as a function of parents' income. The bottom right panel plots the difference in the fraction of children with income above the 90th percentile as a function of parents' income. In each panel, the dashed lines are 95% confidence bands that cover the entire curve with fixed probability. These are calculated using the bootstrap with 500 iterations.
Sources: Panel Study of Income Dynamics, as described in text

References

[1] Chernozhukov, Victor, Iván Fernández-Val, and Alfred Galichon. "Quantile and probability curves without crossing". *Econometrica* 78.3 (2010), pp. 1093–1125.

[2] Van Der Vaart, Aad W and Jon A Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, 1996.